Royal Institute of Technology

ADVANCED MACHINE LEARNING - EM DIRT PUMP & VI

LAST LECTURE

- EM
  - given "structure" and observations find parameters
- EM algorithm for GMM
- Baum-Welch - EM algorithm for training an HMM

THIS LECTURE

- Intuitive derivation of EM
- Derivation of VI
- VI application

RELATIONS BETWEEN LOG-LIKELIHOODS AND Q-TERMS

log-likelihood

Q-term or expected complete log-likelihood (ECLL)

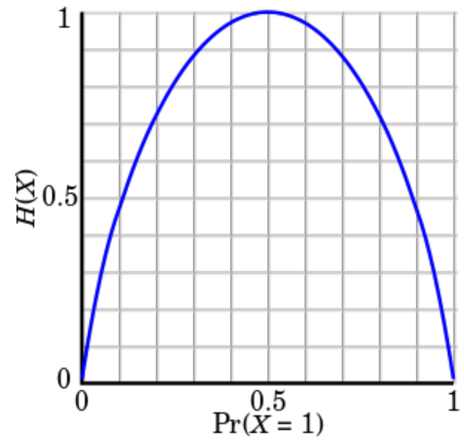
$$Q(\theta, \theta^i) = \sum_n E_{p(Z_n|x_n,\theta^i)} \left[l(\theta; Z_n, x_n) \right]$$

Theorem: by increasing the ECLL (Q-term), we increase the likelihood.

The ECLL may not increase in every step!

USEFUL INFORMATION THEORY CONCEPTS

Entropy



$$\mathrm{H}(p) = -\int p(x)\log p(x)\mathrm{d}x$$

Use Kullback-Leibler (KL) “distance”

$$\mathrm{KL}(p\,||\,q) = \int p(x)\log \frac{p(x)}{q(x)}\mathrm{d}x$$

THE PUMP

$$\log p(X \mid \theta^i) = \mathscr{L}(q(Z), \theta^i) + \mathrm{KL}(q(Z) \,||\, p(Z \mid X, \theta^i))$$

THE PUMP

Although negative

$$\log p(X\,|\,\theta^i) = \mathscr{L}(q(Z), \theta^i) + \mathrm{KL}(q(Z)\,||\,p(Z\,|\,X, \theta^i))$$

THE PUMP

$$\log p(X \mid \theta^i) = \mathscr{L}(q(Z), \theta^i) + \mathrm{KL}(q(Z) \,||\, p(Z \mid X, \theta^i))$$

THE PUMP

$$\log p(X \mid \theta^i) = \mathscr{L}(q(Z), \theta^i) + \mathrm{KL}(q(Z) \,||\, p(Z \mid X, \theta^i))$$

THE PUMP

$$\log p(X \mid \theta^i) = \mathscr{L}(q(Z), \theta^i) + \mathrm{KL}(q(Z) \,||\, p(Z \mid X, \theta^i))$$

$q(z) = p(Z \mid X, \theta^i)$ makes KL become 0

THE PUMP

$=$

$\mathscr{L}(q(Z), \theta^i)$

$q(Z) = p(Z \mid X, \theta^i)$

Hardwire θ^i in $q(Z)$

$$\log p(X \mid \theta^i) = \mathscr{L}(q(Z), \theta^i)$$

THE PUMP

 $=$ 

$$\mathscr{L}(q(Z), \theta^i)$$

$$q(Z) = p(Z|X, \theta^i)$$

$$\log p(X|\theta^i) = \mathscr{L}(q(Z), \theta^i)$$

For any θ,

$$\log p(X|\theta) = \mathscr{L}(q(Z), \theta) + \mathrm{KL}(q(Z)||p(Z|X, \theta))$$

THE PUMP

$=$

$\mathscr{L}(q(Z), \theta^i)$

$$\log p(X \mid \theta^i) = \mathscr{L}(q(Z), \theta^i)$$

For max θ $\quad \geq \qquad \Leftarrow \qquad \geq$

$$\log p(X \mid \theta) = \mathscr{L}(q(Z), \theta) + \mathrm{KL}(q(Z) \,||\, p(Z \mid X, \theta))$$

THE PUMP

=

Depends on $q(z)$

$$\mathscr{L}(q(Z), \theta) =$$

$$E_{q(Z)}[\log \frac{p(X, Z \,|\, \theta)}{q(Z)}] = E_{q(Z)}[\log p(X, Z \,|\, \theta)] + C$$

THE PUMP

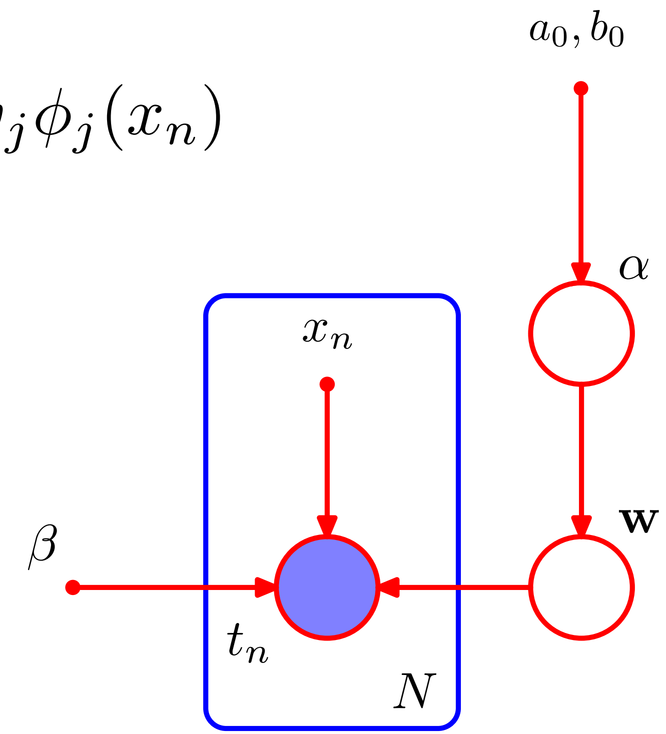
* Initialize θ^0
* Iterate
    - Min KL by setting $q(Z) = p(Z|x, \theta^i)$ so $\log p(x|\theta^i) = \mathscr{L}(q(Z), \theta^i)$
    - Max $\mathscr{L}(q(Z), \theta)$ w.r.t the θ, notice θ^i is “locked” in $q(Z)$, i.e., ECLL
        - $p(x|\theta) > p(x|\theta^i)$ and KL may increase in the eq for this new θ
    - Set θ^{i+1} to θ

BAYESIAN LINEAR REGRESSION

$$x = (x_1, \ldots, x_N)$$

$$t = (t_1, \ldots, t_N)$$

$$t_n \approx w^T \phi(x_n) = \sum_{j=0}^{M-1} w_j \phi_j(x_n)$$



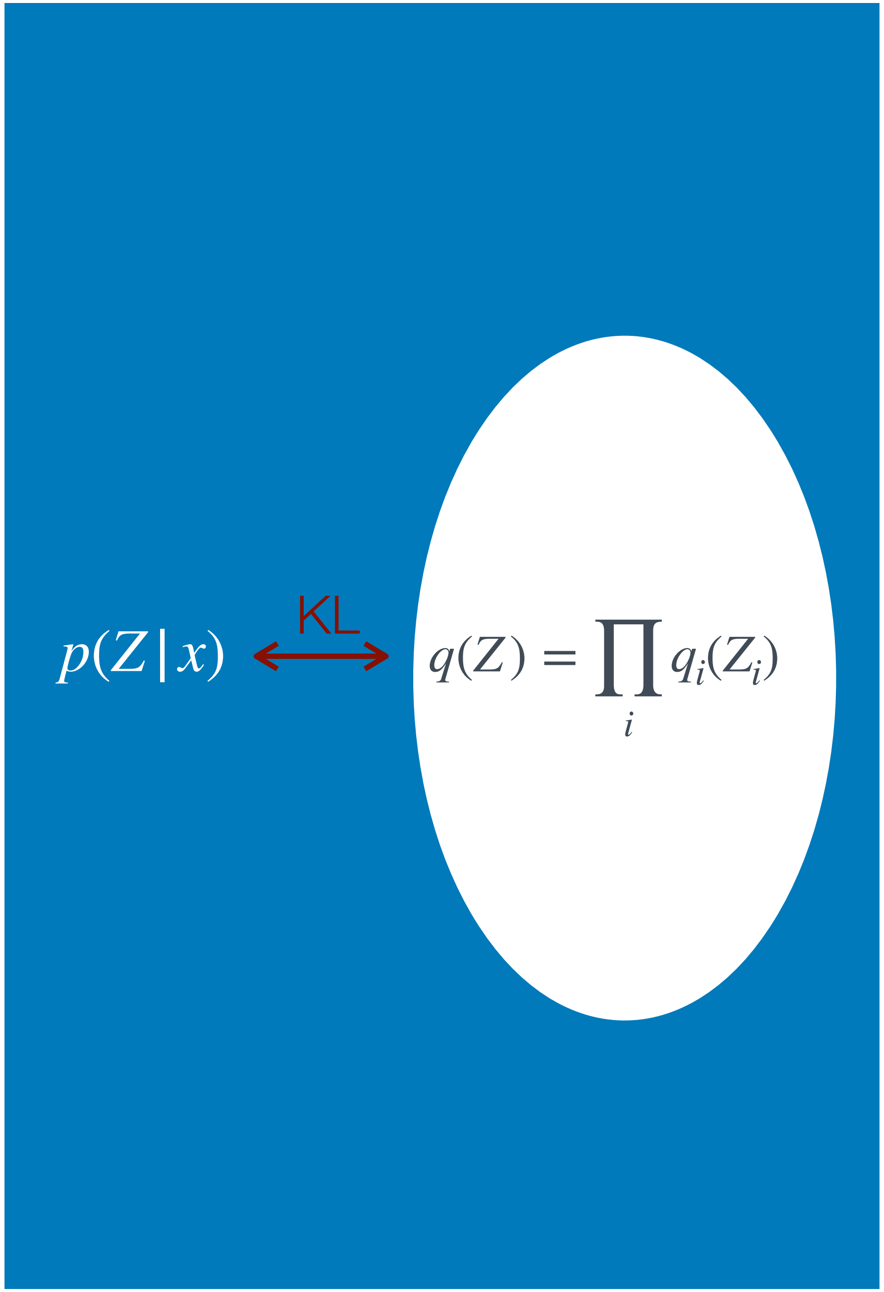
INFERRING POSTERIOR

$$p(t_{N+1}|x_{N+1}, t, x)$$
$$= \int p(t_{N+1}, w|x_{N+1}, t, x)\mathrm{d}w$$
$$= \int p(t_{N+1}|w, x_{N+1})p(w|t, x)\mathrm{d}w$$

- We want to
  - Infer the posterior – Sampling: MCMC, SMC, Gibbs; Deterministic: VI
  - Compute the integral

VARIATIONAL INFERENCE

- Approximate posterior with distribution that is
  - Often gives good approximation
  - Implicitly defined through independence assumption
  - Computationally tractable
  - Obtained by iteratively applying so-called update equations



VI L

$$\mathcal{L}(q(Z)) = \sum_{Z} q(Z) \log \frac{p(X, Z)}{q(Z)}$$

$$\mathrm{KL}(q(Z) \,||\, p(Z \,|\, X)) = \sum_{Z} q(Z) \log \frac{q(Z)}{p(Z \,|\, X)}$$

So,

$$\log p(X) = \mathcal{L}(q(Z)) + \mathrm{KL}(q(Z) \,||\, p(Z \,|\, X))$$

WHICH DISTRIBUTION?

Density up to multiplicative constant

$$Ce^{a_1X^2+a_2X+a_3} \qquad CX^{a_1}e^{a_2X}$$

WHICH DISTRIBUTION?

Density up to multiplicative constant

$$Ce^{-\tau X^2/2+\tau\mu X+C'} \qquad CX^{a-1}e^{-bX}$$

Log density up to multiplicative constant

$$-\tau X^2/2 + \tau\mu X + C' \qquad (a-1)\log X - bX$$

THE END

THE PUMP

$=$

$\mathscr{L}(q(Z), \theta^i)$

$$\boxed{q(Z) = p(Z \mid X, \theta^i)}$$

Depends on $q(z)$

$$\mathscr{L}(q(Z), \theta) =$$

$$E_{q(Z)}[\log \frac{p(X, Z \mid \theta)}{q(Z)}] = E_{\boxed{q(Z)}}[\log p(X, Z \mid \theta)] + C$$